

Prediction of 2024 US election ...*

Colin Sihan Yang Lexun Yu Siddharth Gowda

November 4, 2024

We forecast the winner of the 2024 US presidential election using “poll-of-polls” by building a linear model.

1 Introduction

Election result forecasting has become an essential tool for analysts in political science and the public to predict the outcome of democratic process, such as the presidential election in the United States. Traditionally, individual polls have been used as a snapshot of voter sentiment, but they only reflect temporary changes in the performance of contestants, instead of a precise estimation of the election result. As discussed by Pasek (2015) and Blumenthal (2014), the aggregation of multiple polls, or “poll-of-polls,” has become a popular technique to reduce individual survey errors and provide more accurate election forecasts. However, the traditional poll aggregation does not reflect dynamics of an election, especially with real-time changes and the introduction of new data. This creates a gap for a more adaptable model to predict the election result based on both polling data and additional variables, such as historical data and economic indicators.

This paper fills the gap by building a hybrid election forecasting model following the strategies mentioned by Pasek (2015). As Pasek (2015) described in their article, aggregation involves determining which surveys are worth including, as well as selecting, combining and averaging results from multiple polls to reduce individual biases and errors. Prediction modeling adds other data to the model that predicts election outcomes based on current dynamics. Hybrid models like the Bayesian approach incorporates prior beliefs based on historical data or expert knowledge and new evidence like economic updates to dynamically adjust the forecast as the campaign progresses.

In this paper, we aim to predict the 2024 us election result with the hybrid election forecasting model. We incorporate aggregation by filtering the polls on FiveThirtyEight (2024) by

*Code and data are available at: <https://github.com/yulexun/uselection>.

numeric grade that indicates pollster's reliability, prediction that incorporates social and economic indicators including unemployment rates and abortion rates, and hybrid approaches that leverages Bayesian techniques which combines historical data such as the 2016 election data, allowing for a dynamic prediction of the U.S. presidential election.

The estimand for this research paper is the predicted support percentages for Kamala Harris and Donald Trump. The prediction is based on quantifying various polling factors, including sample size, poll scores, and transparency scores, which are used as predictors.

The results of this model indicate a more stable and accurate forecast compared to traditional aggregation methods alone, [update this ...]

The remainder of this paper is structured as follows: [update this ...]

Appendix

2 Additional data details

3 Model details

3.1 Posterior predictive check

In [?@fig-ppcheckandposteriorvsprior-1](#) we implement a posterior predictive check. This shows...

In [?@fig-ppcheckandposteriorvsprior-2](#) we compare the posterior with the prior. This shows...

3.2 Diagnostics

[?@fig-stanareyouokay-1](#) is a trace plot. It shows... This suggests...

[?@fig-stanareyouokay-2](#) is a Rhat plot. It shows... This suggests...

4 FiveThirtyEight Licenses

[FiveThirtyEight's data sets](#) are used and modified by us under the [Creative Commons Attribution 4.0 International License](#).

5 Trump Voter Prediction Model

The multiple linear regression model (MLR) for Donald Trump will use the same variables, formula, and Bayesian approach as the one for Harris. Likewise, the Trump dataset is also split into training and testing data. The model outputs are below.

Table 1: Head of Bayesian Model Result Summary

Actual_PCT	Predicted_PCT	Lower_CI	Upper_CI
50.7	49.88393	45.11419	54.48807
50.8	49.88235	45.02452	54.71155
49.2	50.94592	46.16608	55.95595
49.2	50.95829	46.24528	55.65467
49.5	50.66132	45.87302	55.27014
50.1	50.63475	45.81683	55.40036
49.6	48.90703	44.33419	53.72756
50.2	48.92830	44.10538	53.62245
48.4	49.10643	43.99543	54.06172
48.3	49.10282	44.39996	53.80928

Figure 1: MLR Trump Model Accounts For A Large Amount of Variability in Voter Percentage

From **?@fig-model-trumpmlr** it is clear that the model accounts for large amount of the variance in Trump’s voter percentage as the data points appear close to the prediction (red line). Furthermore, the distance between the prediction and the actual values do not appear to follow a pattern, suggesting that the error is due to randomness and not model bias.

Based on Figure 3, the test data predictions are also close to the actual values. This suggests that model can generalize to outside data. Similarly, the distance between the prediction and the actual values do not appear to follow a pattern, suggesting that the error is due to randomness and not model bias.

Based on Figure 4, the model expects Trump to win slightly less than 45% of the popular vote and the 95% confidence interval ranges from around 42.5% to 47.2%. This confidence interval is slightly larger than the model for Harris, implying that the Trump polling data might be less reliable.

6 Election Prediction

Our prediction process consists of two primary components. First, we develop models for both Trump and Harris based on the variables outlined in Section **?@sec-PLACEHOLDER**. This

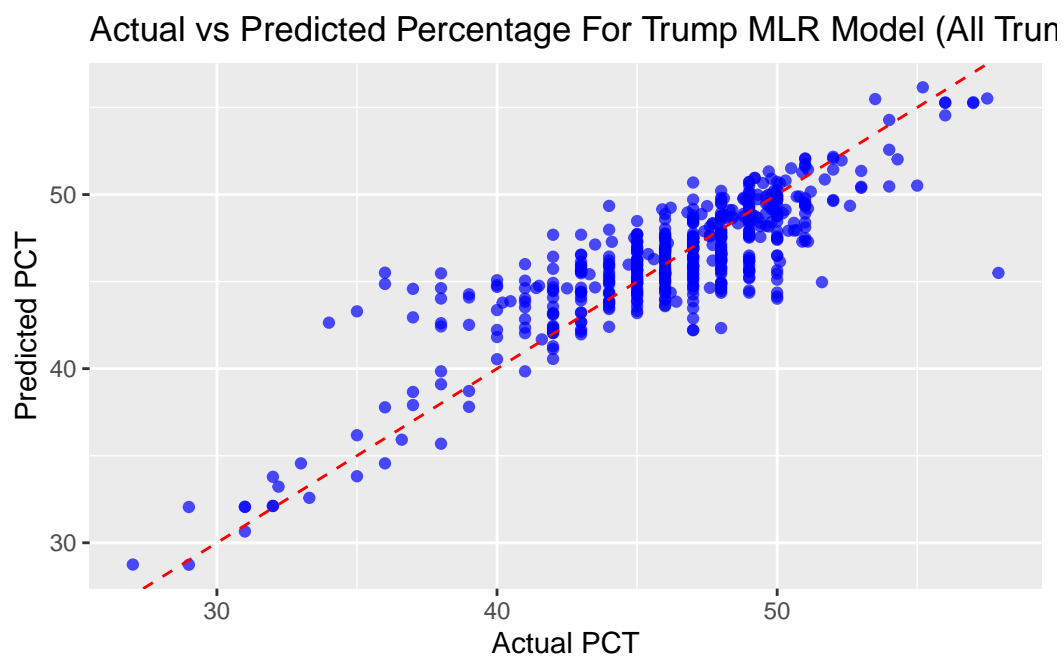


Figure 2: MLR Trump Model Accounts For A Large Amount of Variability in Voter Percentage

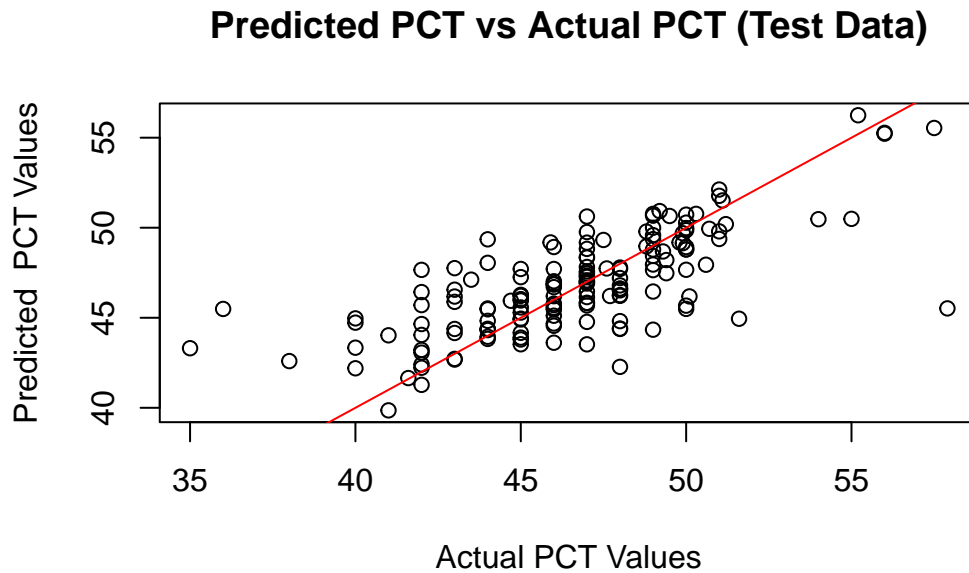


Figure 3: MLR Trump Model does not Appear to Overfit

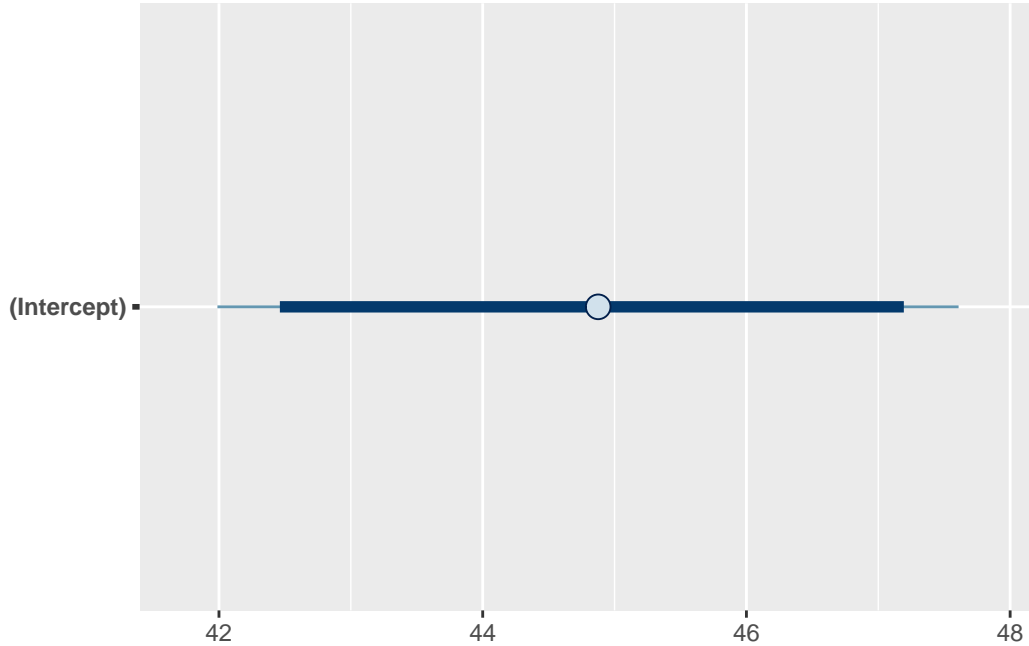


Figure 4: Donald Trump Expected to Recieve Approximatley 45% of the Vote

involves partitioning the dataset into training and testing subsets. Next, we further divide the testing dataset into swing states and other competitive races. We then input this test data into the respective models to generate predictions. By averaging these predictions, we can calculate the expected voter percentage for each candidate in each state. The candidate with the higher percentage is deemed the winner for that state.

We generated predictions for the following states: Arizona, Nevada, Georgia, Pennsylvania, Michigan, Minnesota, Wisconsin, Florida, Texas, Maine CD-2, Nebraska CD-2, New Hampshire, Ohio, Virginia, North Carolina, and Iowa. Winners for other states were determined based on historical trends and predictions from sources like (**cnn?**). Most states without predictions are strongly Republican or Democratic, so their absence is not expected to significantly impact prediction validity.

Table 2: Kamala Harris Wins Most of the Swing States

State	Harris Predicted Percentage	Trump Predicted Percentage	State Winner
Arizona	46.60627	49.20840	Trump
Florida	42.86374	50.74617	Trump
Georgia	47.21810	48.86554	Trump
Iowa	48.56302	43.52854	Harris
Maine CD-2	47.30270	49.28971	Trump

Table 2: Kamala Harris Wins Most of the Swing States

State		Trump Predicted Percentage		State Winner
Michigan	47.59810	46.99337		Harris
Minnesota	48.62212	43.66140		Harris
Nebraska CD-2	49.95245	42.15686		Harris
Nevada	49.35471	47.11387		Harris
New Hampshire	50.78576	42.63839		Harris
North Carolina	48.68842	47.73255		Harris
Ohio	43.96915	50.97722		Trump
Pennsylvania	48.20624	47.17918		Harris
Texas	44.97007	50.26199		Trump
Virginia	49.20564	43.19202		Harris
Wisconsin	48.44502	46.45388		Harris

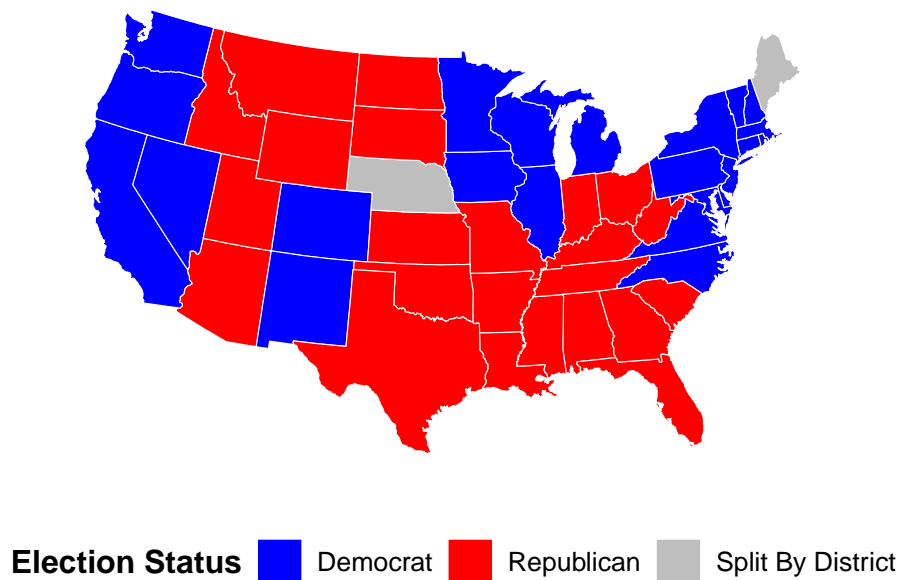


Figure 5: Kamala Harris is Predicated to be the 47th President of the United States

According to Figure ?@fig-currentstateofelection, Kamala Harris is predicted to be the 47th President of the United States. There are also a few states with predictions not visible in the map, we will describe those predicts them below

In Maine, Harris is projected to win the state's overall delegates and District 1, while District 2 is expected to go to Trump. In Nebraska, Trump is expected to win the state's delegates

along with Districts 1 and 3, while Harris is predicted to win District 2. Additionally, Trump is projected to win Alaska, and Harris is expected to win Hawaii.

Overall, the predictions indicate that Harris will receive 298 delegates, while Trump will receive 240 delegates.

References

- Blumenthal, Mark. 2014. “Polls, Forecasts, and Aggregators.” *PS: Political Science and Politics* 47 (2): 297–300. <http://www.jstor.org/stable/43284537>.
- FiveThirtyEight. 2024. “Our Data.” *FiveThirtyEight*. <https://data.fivethirtyeight.com>.
- Pasek, Josh. 2015. “THE POLLS–REVIEW: PREDICTING ELECTIONS: CONSIDERING TOOLS TO POOL THE POLLS.” *The Public Opinion Quarterly* 79 (2): 594–619. <http://www.jstor.org/stable/24546379>.